

The locus and nature of semantic congruity in symbolic comparison: Evidence from the Stroop effect

SAMUEL SHAKI

Bar-Ilan University, Ramat-Gan, Israel

and

DANIEL ALGOM

Tel-Aviv University, Ramat-Aviv, Israel

Pictures of animals with names of animals printed within the pictures were presented for comparative judgments of size based on either the pictures or the names. The picture–word compounds were compared faster with picture than with word as the relevant dimension. The comparisons of pictures were free of interference from the irrelevant names, but the comparisons of names suffered considerable Stroop interference from the irrelevant pictures. Large effects of semantic congruity characterized the comparisons of both pictures and words. Stroop congruity and semantic congruity did not interact even for comparison of words in which both were present, leading instead to additive effects. The results support theories that (1) place semantic congruity in the decision stage and (2) minimize the role of semantic processing as the basis of the semantic congruity effect.

A common cognitive process in everyday life consists of comparing two stimuli and deciding which of the stimuli is larger (or smaller) on the dimension of interest. A remarkable phenomenon associated with such comparisons is the *semantic congruity effect*. When shown a pair of large stimuli, people are faster at selecting the larger of the two; conversely, when shown a pair of small stimuli, people are faster at selecting the smaller of the two. The interaction between the relative magnitude of the stimuli (large or small) and the direction of the comparison (dictated by the instructions “choose larger” or “choose smaller”) defines the congruity effect. Note that deciding which member of a pair of stimuli is larger is logically equivalent to deciding which of the two is smaller; hence, the behavior of a purely logical or a purely perceptual system is not expected to be affected by mere wording of the required comparison. The fact that human observers *are* affected by such wording attests to a confluence of linguistic and perceptual processes in comparative judgment.

Two issues have been salient in the lively theoretical debate engendered by the discovery of the congruity ef-

fect. The first relates to its locus. In various theories, congruity has been assumed to occur prior to the comparative decision (e.g., Duncan & McFarland, 1980; Marschark & Paivio, 1979, 1981; see also Shoben, Sailor, & Wang, 1989), concurrently with the decision (Holyoak, 1978; Jamieson & Petrusic, 1975; Petrusic, 1992), or after the comparative decision (Banks, 1977; Cech & Shoben, 1985; Cech, Shoben, & Love, 1990; see also Cech, 1995). A second and related issue concerns the nature of the comparison process producing congruity. Virtually all theorists agree that the comparative used (“choose larger” or “choose smaller”) is, in some way, semantically processed, but the theorists differ on whether the stimuli themselves are so processed. Propositional or semantic coding models (e.g., Banks, 1977; Banks & Flora, 1977; Banks, White, Stur-gill, & Mermelstein, 1983) assert that they are, whereas analogue (Moyer & Dumais, 1978; Moyer & Landauer, 1967), imagery (e.g., Marschark & Paivio, 1979; Paivio, 1975), reference point (Holyoak, 1978; Marks, 1972), and evidence accrual (Petrusic, 1992) models posit that they are not.

In the present study, we sought to elucidate the locus of the congruity effect by presenting picture–word compounds for comparison on each of the individual dimensions. In this version of Stroop stimuli (Smith & Magee, 1980), names of animals are embedded within pictures of animals as in Figure 1. Matched values on the two dimensions defined congruent stimuli (e.g., the name *cat* embedded within the picture of a cat), and conflicting values defined incongruent stimuli (e.g., the name *cat* embedded within the picture of an elephant). The task for the participant was to decide which member of a pair of stimuli re-

This research was conducted by S.S. under the supervision of D.A. as part of the requirements for an MA degree at Bar-Ilan University. Preliminary results of this study have been presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, November 1997. We thank Craig Leth-Steenen and two anonymous reviewers for insightful comments on an earlier draft. Correspondence concerning this article should be addressed to D. Algom, Department of Psychology, Tel-Aviv University, Ramat-Aviv 69978, Israel (e-mail: algomd@freud.tau.ac.il).

Accepted by previous editorial team

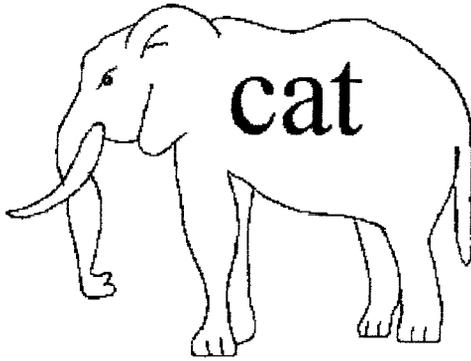


Figure 1. Sample stimulus member used in comparative judgments of picture–word compounds. The compound is an incongruent stimulus because the word does not name the picture.

ferred to the larger (or smaller) animal on the *designated* dimension (e.g., picture), while ignoring differences on the irrelevant dimension (name). Stroop congruity was gauged by the difference in speed and accuracy in comparing congruent and incongruent pairs of stimuli. We were particularly concerned with the relationship between Stroop congruity and semantic congruity. Does semantic congruity differ for Stroop-congruent and Stroop-incongruent pairs of stimuli? In general, do the two effects interact in determining the outcome of stimulus comparison?

A recent study by Cech (1995) may serve as the point of departure for assessing the theoretical significance of our procedure. Applying additive factors methodology (Sternberg, 1969), Cech degraded his stimuli by combining them with dot masks of varying densities or by reducing their luminosity. The manipulations of stimulus quality were effective in that degraded stimuli were compared longer and less accurately than intact stimuli. The major theoretical finding that emerged from Cech's experiments was that stimulus quality did *not* interact with congruity. Impaired performance with degraded stimuli did not entail larger effects of semantic congruity than those obtained with intact or less severely degraded stimuli. Stimulus quality in Cech's experiments was additive with congruity; moving each data point a constant duration sufficed to eliminate the effect of degradation. Essentially the same results were obtained by Banks, Clark, and Lucy (1975; see also Banks, 1977; but see Duncan & McFarland, 1980, and Banks et al., 1983, for somewhat different results and interpretations).

Cech (1995) interpreted the finding of additivity to provide support for multistage theories of stimulus comparison. In that class of models, the encoding stage is assumed to occur entirely prior to the stage containing semantic congruity. Because the congruity effect arises in a later stage, it does not interact with any factor (e.g., stimulus quality, discriminability) influencing the encoding process. Previously, Banks (1977; Banks et al., 1975; see also Banks et al., 1983) offered the same account for data in which stimulus discriminability and congruity led to additive effects. In

models that assume congruity to operate during encoding, the congruity effect ought to interact with stimulus quality or discriminability. Because the results failed to support this prediction, Cech rejected the latter class of models as a viable explanation of stimulus comparison and congruity.

The picture–word stimuli used in this study may be conceived of as a novel species of degraded stimuli. In previous experiments, stimuli were degraded *physically* by a wealth of means, including diagonal line-grids (Duncan & McFarland, 1980), rub-on letters (Banks et al., 1983), dot masks and reduced luminosity (Cech, 1995), or decreased discriminability (Banks et al., 1975). In the present study, by contrast, degradation was *semantic*. Perceptually, both the names and the pictures presented were eminently legible (see Figure 1). As a result, any cost to performance incurred by incongruent pairs of stimuli—the Stroop effect—would have to emanate from semantic interference. Of course, if the participant were able to ignore the pictures when comparing names, no Stroop congruity would emerge. Alternatively, if the participant could not ignore the pictures when comparing names, impaired performance with incongruent pairs of stimuli (i.e., Stroop congruity) would ensue. Note, though, that the impairment to performance would derive wholly from the semantic processing of the irrelevant dimension, not from any form of physical degradation.

Presenting Stroop-like stimuli for comparison introduces semantic processing into the encoding stage. The encoding stage entails those processes that are needed for identifying the pertinent stimuli. For symbolic comparisons, encoding further entails retrieval of the respective values from the referent dimensions. Clearly, comparing a pair of stimuli requires that the individual values are first perceived. These elementary processes of identification or valuation typically involve sensory and/or perceptual processes. Our procedure imported linguistic influences into encoding, a stage that otherwise has been considered to be largely perceptual. By using both forms of the comparative (“choose larger” and “choose smaller”), a routine procedure, we engendered effects of semantic congruity. Putatively, both Stroop and semantic congruity effects entail semantic processes; our procedure thus offered many opportunities for interaction—if semantic congruity occurs during encoding. If semantic congruity shares in any of the resources involved in encoding, Stroop and congruity effects ought to coalesce, the outcome of the comparison depending on their interaction. Alternatively, if congruity is completely separate from encoding, no interaction of the two effects would be expected to emerge. In that case, the semantic variations produced by Stroop effects at encoding should be additive with those produced by congruity effects at comparison. Following additive factors logic, our procedure provided a strong test of the locus of semantic congruity (in particular, of whether or not it is separate from encoding).

Notably, within the class of semantic models, the congruity effect itself has been construed to be a form of Stroop interference (Banks & Root, 1979; Leth-Steensen

& Marley, 2000). Banks and Root have suggested a *semantic interference model* in which the congruity effect derives from conflicting codes generated in parallel. The model relies “on a form of competition among processing codes much like Stroop interference” (Banks & Root, 1979, p. 140). In a comparison task, the magnitude of large stimuli is an irrelevant attribute that generates the implicit code “large,” which interferes with processing of the “choose smaller” instruction but facilitates processing of the “choose larger” instruction. In Leth-Steensen and Marley’s (2000) theory, the relative magnitude of the stimuli selectively influences competing codes generated by the two instructions. Banks and Root’s *semantic interference model* is a single-stage model unlike Banks’s (1977; see also Cech & Shoben, 1985) more influential *semantic coding model*. The latter model entails code manipulations in a set of serial processing stages, with stimuli (and instructions) ascribed codes *before* the comparative decision is made. The semantic interference model has not yet been tested. Our study provides the first test of this theory: If semantic congruity comprises an instance of Stroop interference, it ought to interact with another instance of Stroop interference, one generated concurrently by the experimental stimuli.

Finally, presentation of Stroop-like stimuli may also serve to test the class of analogue theories in which stimuli are represented as points on an underlying continuum, the end anchors of which are labeled reference points. For instance, Holyoak (1978) and Jamieson and Petrusic (1975) have discussed a model in which comparative reaction time is a function of the ratio of the distances of the stimuli from the end anchor activated by the instructions (“choose larger” or “choose smaller”). For a fixed interstimulus distance, the ratio differs according to the end anchor from which the distances are calculated, hence the congruity effect. Evidence accrual theories (Petrusic, 1992) posit semantic congruity to occur within each glimpse as the pair of stimuli is sampled continuously over time on a given trial. The point to note is that in analogue theories, the stimuli do not undergo semantic processing (i.e., they are not ascribed labels such as “large” or “small”); semantic processing of the instructions is also minimal and amounts to directing the observer to the appropriate end anchor. Its name notwithstanding, the semantic congruity effect does not derive from semantic interference (or facilitation) between the stimuli and the instructions. Following the logic of analogue and evidence accrual models, the semantic processes producing the Stroop effect should not interact with the essentially nonsemantic processes producing the semantic congruity effect. This prediction is tested in our study.

Apart from studying semantic congruity, the main focus of the present research, we also examined two other effects known to characterize comparative judgments, the distance and end effects. The distance effect concerns the separation of the pair of stimuli on the referent continuum: The greater the separation, the smaller the reaction time (RT) needed to decide which of the two stimuli is the larger or smaller. The end effect registers the fact that pairs of

stimuli containing an extreme value of the scale as a member are compared fastest. We tested the influence of the Stroop effect on these aspects of the comparative process too.

Virtually all investigators of comparative judgment have espoused a unidimensional design. The stimuli presented for comparison have differed solely on the dimension of interest. This contrasts with the widespread use of multidimensional stimuli in the allied domain of comparisons with *same-different* responses (see Farell, 1985, for a review). In the present study, we presented multidimensional stimuli for comparative judgments of magnitude on an individual dimension. Participants compared stimuli on one dimension but had to ignore systematic variation along an irrelevant dimension. The failure of selective attention to the relevant dimension was gauged by Stroop congruity. By presenting Stroop-like stimuli, we were able to produce results that place constraints on theories of stimulus comparison.

EXPERIMENT 1

Method

Participants. Twenty young men and women, recruited from the Bar-Ilan University community, were paid to participate in two experimental sessions.

Stimuli and Apparatus. The stimulus set consisted of the Hebrew names for *elephant*, *cow*, *cat*, and *fly*, and outline drawings of the same animals. The animal names (all three-to-four-letter words in Hebrew) appeared embedded within the pictures of the animals as in Figure 1. In *congruent pairs*, the words named the pictures of the animals. For instance, in one congruent pair, the stimulus on the left comprised the name and picture of a cow, and the stimulus on the right comprised the name and picture of an elephant. In *incongruent pairs*, at least one member of the pair consisted of a conflicting picture-word stimulus (i.e., the picture and the name did not refer to the same animal). We also used one-dimensional or *neutral pairs* for comparisons of both names and pictures. For the former, the name was embedded within an outline drawing of a “cloud,” about the same size as the pictures. For pictures, the neutral stimuli entailed pictures of animals enwrapping three-member-strings of the Hebrew letter *Quoph*, a letter that did not appear in any of the animal names.

A given pair appeared at both spatial arrangements (i.e., a member appeared once on the left and once on the right) and was shown with each of the two comparative questions, “Which animal is larger?” and “Which animal is smaller?” Each combination of stimuli, relative location, and instruction was presented twice. The participants performed a total of 624 experimental trials (78 stimuli \times 2 spatial arrangements \times 2 instructions \times 2 repetitions), including 48 congruent, 528 incongruent, and 48 neutral pairs of stimuli. To understand the partition, consider all the incongruent combinations for a given pair of words. There were seven combinations that involved non-identical pictures and four combinations in which one of the words named its picture but the other did not. Hence, there were 11 incongruent items per each congruent item. The experimental trials were preceded by 160 practice trials that comprised the three types of stimuli in their experimental proportions. The entire series of 784 trials was presented in one block; the participants were not aware of the partition into practice and experimental trials. Order of presentation was random and different for each participant.

The pair of picture-word stimuli appeared at the respective centers of the left and right hemifields on the gray background of an IBM-compatible (PC 486) computer screen. Each picture measured 366 \times 280 pixels on the screen. The word stimuli were generated in Hebrew font Miriam and were displayed at the centers of the respective

pictures. The names were presented within a 200×52 pixel frame (invisible), with care taken to space letters proportionally within a word. The participant sat approximately 80 cm from the center of the screen.

Procedure. The participants were tested individually in a dimly lit room. Each participated in two experimental sessions separated by at least 48 h. In one session, the participants were instructed to attend to the pictures and ignore the names. In another session, they were instructed to attend to the names and ignore the pictures. Half of the participants first performed the picture task, and half first performed the name task. Comparisons were made by pressing either a right- or a left-hand key on the keyboard. The participant's task was to press the key on the same side of the larger (or smaller) member of the pair of items. Because the stimuli were of equal physical size, the necessity of reacting to the *conceptually* larger or smaller member of each pair was emphasized.

The participant initiated the first trial by pressing any key on the computer keyboard. The comparative question followed 0.5 sec after each response. After an additional 0.5 sec, the pair of stimuli appeared while the comparative question remained on the screen. The stimuli and the comparative question were response terminated. Reaction time (RT) was measured in milliseconds with the use of a software timer. The participants were encouraged to respond quickly but accurately. During a session, they could take as many breaks as they deemed necessary, an option that only a few exercised. Each session lasted about 1 h.

The Pearson correlation between RT and error was .26 ($p > .10$) for names and .13 ($p > .10$) for pictures.

Results

A summary of the results appears in the top panel of Figure 2. Mean RT for selecting the larger or smaller member of the pair is plotted as a function of pair type (congruent, incongruent, or neutral) for comparisons of names and pictures. A glimpse at Figure 2 reveals that pictures were compared much faster than names. On average, the speed advantage of pictures amounted to 210.1 msec [$F(1,19) = 21.18, MS_e = 187,557, p < .001$]. Selective attention to pictures was good: Comparison times did not differ appreciably across congruent, incongruent, and neutral pairs of stimuli [$F(2,38) = 1.7, MS_e = 2,802, p > .10$]. Hence, the Stroop effect was absent for pictures as the criterial dimension. The participants could attend to the pictures without suffering intrusions from the irrelevant names. By contrast, selective attention failed for the slower dimension of name: The participants could not attend to the names without suffering intrusions from the irrelevant pictures. A sizeable Stroop effect was obtained for name as the criterial dimension, with congruent pairs compared 80.5 msec faster than incongruent pairs [$F(2,38) = 20.81, MS_e = 4,661, p < .001$]. The interaction of criterial dimension and pair type [$F(2,38) = 12.82, MS_e = 3,626, p < .001$] confirmed the presence of a Stroop effect for comparisons of names but its absence from comparisons of pictures. Our participants could ignore irrelevant names when they were comparing pictures, but they were unable to ignore irrelevant pictures when they were comparing names.

The Stroop effect obtained for names is composed of Stroop *facilitation* (the difference in RT between congruent and neutral pairs) and Stroop *interference* (the complementary difference in RT between neutral and in-

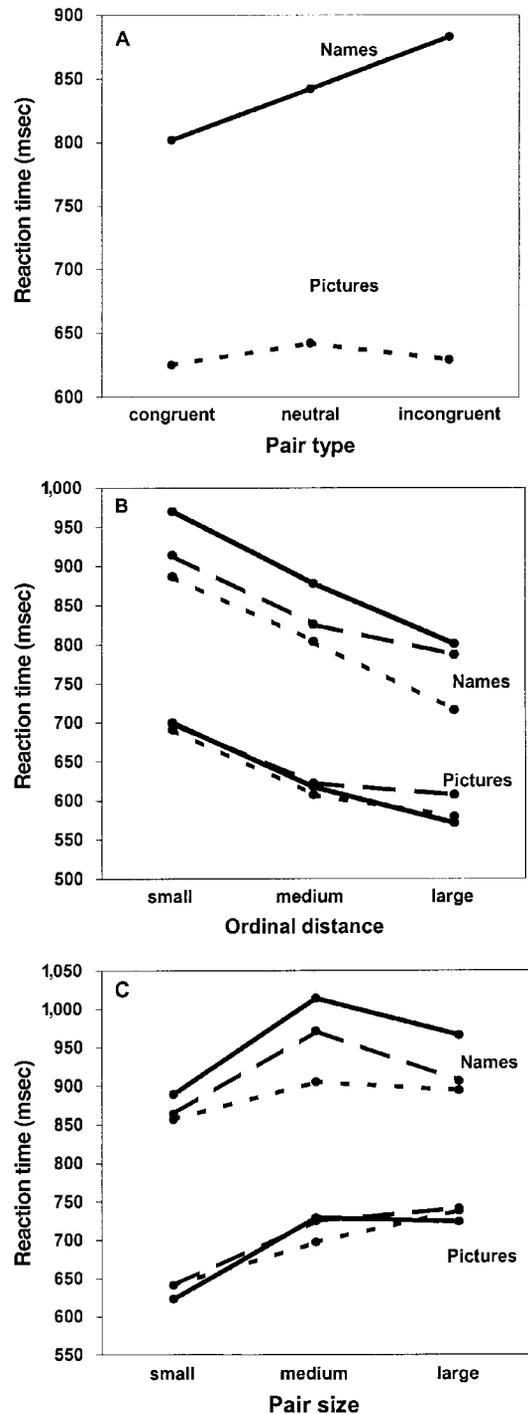


Figure 2. (A) Results of Experiment 1 showing both faster processing of pictures than of names and a Stroop effect for comparison of names but the absence of a Stroop effect for comparison of pictures. (B) The distance effect: Comparative reaction time declines as the difference in referent size increases. The distance effect is about the same magnitude for names and pictures and, within each dimension, for congruent, neutral, and incongruent pairs of stimuli. Apparent again is the faster processing of pictures and the confinement of the Stroop effect for comparison of names. (C) The end effect: Reaction time declines for pairs that include an extreme value of the scale as a member (names) or for the smallest pair (pictures).

congruent pairs of stimuli). In the present data, facilitation (39.72 msec) and interference (40.72 msec) were approximately equal in magnitude [$F(1,19) = 13.0, MS_e = 3,627, p < .001$, and $F(1,19) = 13.72, MS_e = 3,626, p < .001$, respectively]. Comparisons of names were facilitated by congruent pictures and impaired by conflicting pictures relative to neutral stimuli that entailed no pictures. For the complementary comparisons of pictures, the participants neither reaped gain nor suffered interference from congruent or conflicting names.

The speedier performance with pictures was also more accurate: Overall, the participants made fewer errors with pictures (2.4%) than with names (3.5%) [$t(38) = -1.813, p < .038$]. Therefore, a superior comparison performance with pictures was exhibited through both speed and accuracy. Within each dimension, Stroop effects of 1% for both pictures and names were not statistically significant.

In the middle panel of Figure 2, we replotted the RT data as a function of the ordinal distance separating the two animals in a pair. The three curves for each dimension describe the responses for congruent, neutral, and incongruent pairs of stimuli. For comparisons of pictures, the classic *distance effect* obtained (Banks, 1977), whereby the time needed to decide which animal was larger or smaller decreased as the ordinal separation in size between the animals increased [$F(2,38) = 41.78, MS_e = 4,733, p < .001$]. The same effect of distance characterized the comparisons for all three types of stimuli, with the respective curves coalescing into a single distance function [$F(4,76) = 1.345, MS_e = 1,529, p > .26$]. The absence of an effect for pair type confirms the absence of Stroop congruity for pictures.

A distance effect characterized the comparisons of names as well [$F(2,38) = 50.72, MS_e = 7,228, p < .001$]. Indeed, the same distance effect [$F(2,38) = 2.55, MS_e = 7,001, p > .09$] characterized both types of comparisons, as was revealed by the absence of an interaction between criterial dimension and distance [$F(2,38) = 2.55, MS_e = 7,001, p > .05$]. However, for comparisons of names, performance differed across pair type. It was best for congruent pairs and worst for incongruent pairs, the vertical difference of 80 msec between the respective curves reproducing the Stroop effect calculated on the basis of the data summarized in the top panel of Figure 2. Figure 2 also shows that the separation along the vertical axis of the distance functions was fairly constant, with the visual parallelism supported by an insignificant distance \times stimulus type interaction [$F(4,76) = 1.34, MS_e = 3,253, p > .2$]. The analyses of the distance effects reproduced the main result apparent in the top panel of Figure 2: Comparisons of names were affected by an appreciable Stroop effect, but comparisons of pictures were free of Stroop interference.

The independence of Stroop and distance effects is notable. The Stroop effect (for names) was approximately additive with the distance effect. Our results suggest that the Stroop effect and the distance effect derive from different stages of processing. If the Stroop effect results from the encoding stage, the distance effect must result from a sub-

sequent stage, conceivably from the very stage in which the comparative decision is made.

The error data largely reproduced the RT data vis à vis the distance effects. For both comparison of pictures and names, the participants made fewer errors with pairs entailing large ordinal separation of components as opposed to small separation of components [$F(2,57) = 3.96, p < .05$ for pictures, and $F(2,57) = 13.2, p < .01$ for names].

A subset of the data from the top panel of Figure 2 (pairs separated by a single ordinal unit) is replotted in the bottom panel of Figure 2 for an examination of the end effect. Comparison RT is plotted against stimulus size for names and pictures, separately for congruent, neutral, and incongruent stimuli. Evident in the bottom panel of Figure 2 is the large separation in RT between names and pictures, as large indeed in this subset of the data as it is in the full set of data (top panel). The speed advantage for comparison of pictures over comparison of names amounted to an average of 223 msec [$F(1,19) = 16.18, MS_e = 553,343, p < .007$].

For names, the classic inverted-U-shaped effect obtained: Performance was fastest for the smallest (870.42 msec) and largest (922.52 msec) stimulus pairs and was slowest with pairs of intermediate size (963.92 msec) [$F(2,38) = 11.11, MS_e = 23,699, p < .002$]. A large Stroop effect was again tapped by the vertical separation of the curves for the three types of stimuli. Mean RT was fastest for congruent stimuli (885.66 msec), intermediate for neutral stimuli (914.22 msec), and slowest for incongruent stimuli (956.98 msec) [$F(2,38) = 11.09, MS_e = 13,932, p < .002$].

For comparison of pictures, a monotonic end effect obtained: Performance was fastest for the smallest pairs (635.46 msec), slower for the intermediate pairs (717.52 msec), and slowest for the largest pairs (734.91 msec) [$F(2,38) = 21.76, MS_e = 15,562, p < .001$]. For comparisons of the pictures, the three curves for congruent, neutral, and incongruent pairs coalesced into a single function ($F < 1$). The lack of separation again documents the absence of a Stroop effect in comparison of pictures.

The participants made fewer errors with pictures (3.2%) than with names (4.29%) in this subset of the data, but the difference was not significant statistically [$t(38) = -1.4, p > .05$]. For comparison of both pictures and names, the participants made fewer errors with small pairs than with large pairs, but the difference was statistically significant only for the names [$F(2,38) = 2.14, p > .05$ for pictures, and $F(2,38) = 6.7, p < .003$ for names].

In the left portion of Figure 3, we portray the effects of semantic congruity obtained for comparisons of pictures and names, separately for congruent and incongruent stimuli (we omitted neutral pairs and intermediate size from Figure 3 for sake of legibility, although the analysis includes all of the values). Figure 3 includes the same data used for depicting the end effect—pairs separated by a single ordinal unit of distance. A large difference in overall RT between comparisons of pictures and names is again apparent. The magnitude of the semantic congruity effect

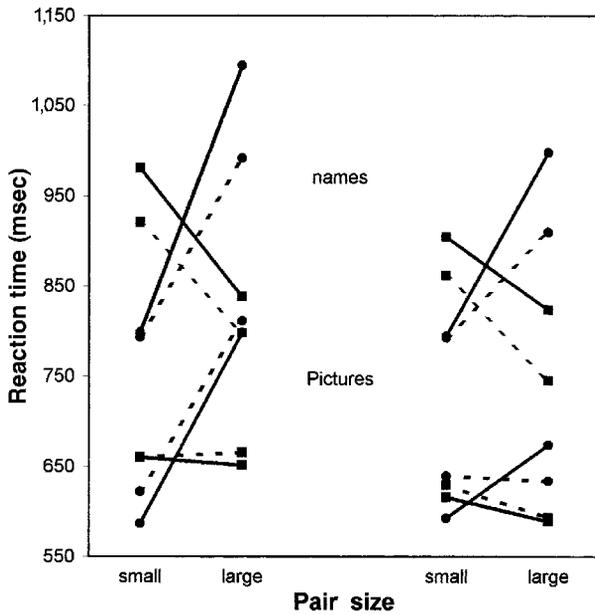


Figure 3. Left side: Semantic congruity effects for pictures and names and, within each dimension, for congruent (broken line) and incongruent (solid line) pairs of stimuli. The results show large crossover effects for both names and pictures. The vertical separation of the curves depicts an appreciable Stroop effect for names, but a minuscule reverse Stroop effect for pictures. The parallelism of the respective curves for congruent and incongruent pairs of stimuli demonstrates the lack of an interaction between Stroop and semantic congruity. The data show the effects for comparisons of pairs entailing adjacent stimuli. Right side: Same as left side, for nonadjacent comparisons entailing pairs separated by two units of ordinal distance.

differed across pictures and names. On average, the size of the semantic congruity effect was 197.5 msec for pictures and 377.8 msec for names (the magnitude of the effect was defined by the average time needed to be added or subtracted in order to align the curves for the two instructions) [$F(2,38) = 5.97$, $MS_e = 20,542$, $p < .006$].¹ The difference likely derives from general speed of responding: The larger the RT, the larger the effect of semantic congruity. A similar decline of approximately 40% has often been observed (Banks & Flora, 1977; Figure 3).

For pictures, appreciable effects of semantic congruity obtained for all three classes of stimuli [$F(2,38) = 6.725$, $MS_e = 13,534$, $p < .003$ for congruent stimuli; $F(2,38) = 10.391$, $MS_e = 11,603$, $p < .003$ for neutral stimuli; and $F(2,38) = 12.117$, $MS_e = 11,063$, $p < .001$ for incongruent stimuli]. The mean RTs for the three types of pairs were 692.36, 703.1, and 692.44 msec, respectively, for congruent, neutral, and incongruent pairs ($F < 1$), yielding a zero Stroop effect. The effects of semantic congruity did not differ across pair type, with the absence of an interaction of size, instruction, and stimulus type [$F(4,76) = 2.496$, $MS_e = 9,340$, $p > .05$] documenting the complete absence of a Stroop effect from these data. Visually, the minuscule vertical separation of the curves for congruent and incongruent stimuli demonstrates the absence of

a Stroop effect from semantic congruity with pictures (actually, a small tendency for a reverse Stroop effect is discernible).

For names, a large effect of semantic congruity was accompanied by an equally substantial Stroop effect. For comparison of names, the overall RTs differed across pair type (means of 885.7, 914.2, and 956.9 msec for congruent, neutral, and incongruent pairs, respectively), yielding a Stroop effect of 71.32 msec [$F(2,38) = 11.10$, $MS_e = 13,933$, $p < .0002$]. Salient in Figure 3 is the vertical separation of incongruent and congruent stimuli in comparisons of names. Effects of semantic congruity obtained for all three types of stimuli [$F(2,38) = 20.48$, $MS_e = 14,962$, $p < .000$ for congruent pairs, $F(2,38) = 23.3$, $MS_e = 14,963$, $p < .000$ for incongruent pairs, and $F(2,38) = 50.38$, $MS_e = 9,494$, $p < .000$ for neutral pairs]. The large crossover interactions obtained for both congruent and incongruent stimuli are evident in Figure 3. However, the Stroop and semantic congruity effects obtained did *not* interact ($F < 1$) and led instead to additive effects. Because of the theoretical importance of this null effect, we calculated power on the basis of advice given by Winer (1971) and assisted by the computer program developed by Faul and Erdfelder (1992; we should note that no consensual routine for calculating the power of three-way interactions like the present one exists in the literature). The power was estimated to be 0.73, characterized as moderate by Cohen (1988). Visually, the crossover effects were approximately equal in size for the congruent and incongruent pairs of stimuli (see Figure 3).

The pattern of data for errors mirrored that obtained for RT. For pictures, the participants made fewer errors when the relative magnitude of the stimuli (large or small) was congruent with the direction of the comparative ("choose larger" or "choose smaller") than vice versa [$F(1,19) = 5.51$, $MS_e = 2.761$, $p < .03$]. For names, the semantic congruity effect was not significant ($F < 1$). We should note that observing congruity effects with an accuracy measure is rare (see Petrusic, 1992, and Petrusic & Baranski, 1989, for discussions of the role of accuracy in comparative judgment).

To recap, comparisons of names and pictures alike were characterized by effects of semantic congruity, yet only the former was also affected by Stroop interference. However, Stroop and semantic congruity effects did not interact even in the case when both were presented (for names). To provide further evidence on the dissociation of semantic congruity and Stroop congruity, we decided to examine another small subset of the data—pairs separated by two units of ordinal distance. In the right-hand part of Figure 3, we present these results (the presentation is tailored after that of the left side). The large difference in overall RT between comparison of pictures (616.53 msec) and names (857.69 msec) is again apparent [$t(38) = -5.476$, $p < .000$], as are the effects of semantic congruity for both pictures [$F(1,19) = 6.03$, $MS_e = 3,892$, $p < .02$] and names [$F(1,19) = 51.13$, $MS_e = 8,024$, $p < .000$]. There was a difference in the size of the congruity effect for

pictures (72.6 msec) and names (286.4 msec) [$F(1,38) = 19.93, MS_e = 5,958, p < .000$]. For comparisons of pictures, incongruent pairs yielded a crossover effect, congruent pairs a funnel effect.²

The feature of focal interest was, again, the relationship between the Stroop and semantic congruity effects. A Stroop effect (73.3 msec) emerged only for names. For comparison of pictures, the RTs did not differ across stimulus type. Stroop and semantic congruity effects did not interact in comparison of names [$F(2,38) = 1.29, MS_e = 12,312, p > .5$], leading again to additive effects. In this subset of the data, too, semantic congruity characterized comparisons of both pictures and names, yet a Stroop effect emerged only for names. With both effects present for names, they influenced symbolic comparisons in an additive rather than an interactive fashion.

Discussion

The most important result to emerge from this experiment is the additivity of Stroop and semantic congruity effects in symbolic comparison. Let us recount the dissociation. First, the effects did not correlate across the sets of comparisons with pictures and words. Semantic congruity characterized performance with both pictures and words, yet the Stroop effect affected only that with words. Second, the effects were additive even when both were present in the data (for words). This result, depicted in Figure 3 (left side), deserves another glimpse.

For words, we calculated a Stroop effect of approximately 70 msec—the amount by which congruent pairs were compared faster than incongruent pairs. This Stroop effect appears as the fairly constant vertical distance separating the curves for congruent and incongruent pairs. In contrast, the semantic congruity effect remained invariant across congruent and incongruent pairs (approximately 380 msec for both congruent and incongruent pairs). Apparent in the graphical display is the similar form of the two crossover interactions. The same crossover interaction of stimulus size and instruction, typical of symbolic comparisons, characterized both the congruent and the incongruent pairs. Statistical analysis supported the observed independence of Stroop and semantic congruity. The two did not interact and led instead to additive effects. Therefore, Stroop congruity and semantic congruity appear to be unrelated phenomena, tapping separate stages of the comparison process.

The distance effect (see Figure 2), too, was additive with the Stroop effect. The distance effect thus results from a different stage of processing than does the Stroop effect. Several lines of evidence converge on the conclusion that the Stroop effect operates in the encoding stage. First, when pictures and words are presented (separately) for comparative judgment, only the former is affected by Stroop congruity (Banks, 1977; Banks & Flora, 1977; Paivio, 1975). Because pictures and words are processed alike in all the other stages of the comparison process, the Stroop effect is relegated by default to the encoding stage. Second, “since the overall picture–word RT difference

takes place in the in the encoding stage, then the interaction between the picture–word variable and the . . . [Stroop] congruity effect indicates that . . . congruity has its effects there too” (Banks & Flora, 1977, p. 289). Third, logic dictates the placement of the Stroop effect in the encoding stage. Comparing things presupposes identifying each on the dimension of interest. Because the Stroop effect influences this elementary process of stimulus specification, it is part and parcel of stimulus encoding. By the same token, the distance effect does not result from the encoding stage. Could the distance effect comprise part of the comparative decision itself (cf. Duncan & McFarland, 1980; Petrusic, 1992)?

The incommensurability of Stroop and semantic congruity, on the one hand, and the incommensurability of Stroop and the distance effect, on the other hand, carry ramifications for theoretical resolution. We defer drawing the pertinent conclusions for the General Discussion section, however, after we have examined the results of a second experiment probing the same phenomena.

In Experiment 1, we used a four-term order so that the analyses of the semantic congruity effect involved the end terms. Expectancy or knowledge of the stimulus structure could have confounded the results. Although we have no evidence for such an influence, we cannot rule it out. To reduce the possibility of strategic responding, in Experiment 2, we used a wider range of stimuli. We espoused a design that approximated the infinite-set version of the comparison task (Banks & Flora, 1977; Paivio, 1975). Each participant was presented with a small sample of the entire stimulus ensemble so that a given pair of stimuli was judged only once by the same participant. The question posed in Experiment 1 was tested again: Do Stroop and semantic congruity exert independent effects on comparative judgment?

EXPERIMENT 2

Experiment 2 was designed as a replication and extension of Experiment 1. It served to answer the following question. Might the additivity of semantic congruity and Stroop effects observed in Experiment 1 imply merely that our stimulus values were too few? To answer this question, in Experiment 2, the comparative judgments involved eight animals, spanning the size range between fly and elephant. Moreover, we presented the pairs of picture–word compounds within the framework of a design that approximated (although did not fully follow) an infinite-set paradigm. In that paradigm, the participant never sees a given item twice. In our design, of the 1,568 unique pairs available, each participant was presented with a random sample of 40 pairs; a given pair of picture–word compounds was presented only once. An associative network based on repeated presentation of the same stimuli is implausible under this design. Confounding by expectations or by knowledge of the stimulus structure are equally implausible. The data generated by applying this design served to validate the results and conclusions obtained in Experi-

ment 1. The separability of Stroop and semantic congruity effects should be revealed again by additive influence on the outcome of the comparative judgment.

Method

Participants. The participants were 32 young men and women associated with the Tel-Aviv University community who volunteered to participate in two short experimental sessions. None had participated in Experiment 1.

Stimuli and Apparatus. The stimulus set consisted of the Hebrew names for *fly, butterfly, bird, cat, dog, lion, cow, and elephant*. The animal names (again, all three- to four-letter words in Hebrew) appeared embedded within the pictures of the animals as in Figure 1. In *congruent pairs*, the words named the pictures of the animals. In *incongruent pairs*, at least one member of the pair consisted of a conflicting picture–word stimulus. We included no one-dimensional or neutral pairs of stimuli in Experiment 2.

Producing all pairwise combinations of the picture–word compounds with both comparatives (“Which animal is larger?” and “Which animal is smaller?”) resulted in 1,568 unique stimulus items. With the use of our computer program, we selected from this population random samples of 40 pairs without replacement, a different random sample for each participant. The only other constraint set on the process of sampling was that each sample was to preserve (conservatively) the proportion of congruent pairs in the population (i.e., .036). This meant that each individual sample contained *at least 2* congruent pairs. The conditions of presentations were the same as those of Experiment 1.

Note that in a truly infinite-set design, the participant never sees the same picture or the same word more than once. This was not the case with our design. Our participants could have been presented with same picture or the same word more than once (virtually all were), but they were not presented with the same pair of picture–word compounds more than once. Hence, the present experiment comprised an approximation to a strictly infinite-set design. The infinite-set design is a taxing method with unidimensional stimuli (hence its sparsity), and it is impracticable with the multidimensional stimuli used in the present study.

Procedure. The participants performed in two short sessions of approximately 10 min each, separated by a break of 20 min. During a session, the participant was presented with 40 pairs of picture–word compounds for comparative judgment. In one session, the participant was instructed to respond to the pictures and ignore the names; in another session, the participant responded to the names, attempting to ignore the pictures. Method of responding and all the other details of the procedure followed those used in Experiment 1.

In the ensuing analyses of variance (ANOVAs), for the unequal sample sizes in some of the analyses we followed the recommendation and routines detailed by Keppel (1973). Rates of errors were low, averaging at 4.83% for pictures and 5.96% for words [$F(1,31) = 1.2, MS_e = 177, p > .05$]. None of the other comparisons involving errors turned out to be statistically significant for either pictures or names. The correlations between RT and errors were .39 ($p > .05$) for pictures and .52 ($p > .05$) for words, indicating an absence of a speed–accuracy tradeoff in this experiment too. Given the unique structure of the present experiment, we discuss the RTs and do not treat the errors further.

Results

Figure 4 (top panel) provides a summary of the results. Pictures were compared faster than were names by 190.6 msec [$F(1,31) = 8.52, MS_e = 125,542, p < .01$]. Stroop congruity was absent from the comparisons of pictures (amounting to 12.4 msec, $F < 1$), but it was substantial, 152.7 msec in comparisons of names [$F(1,31) =$

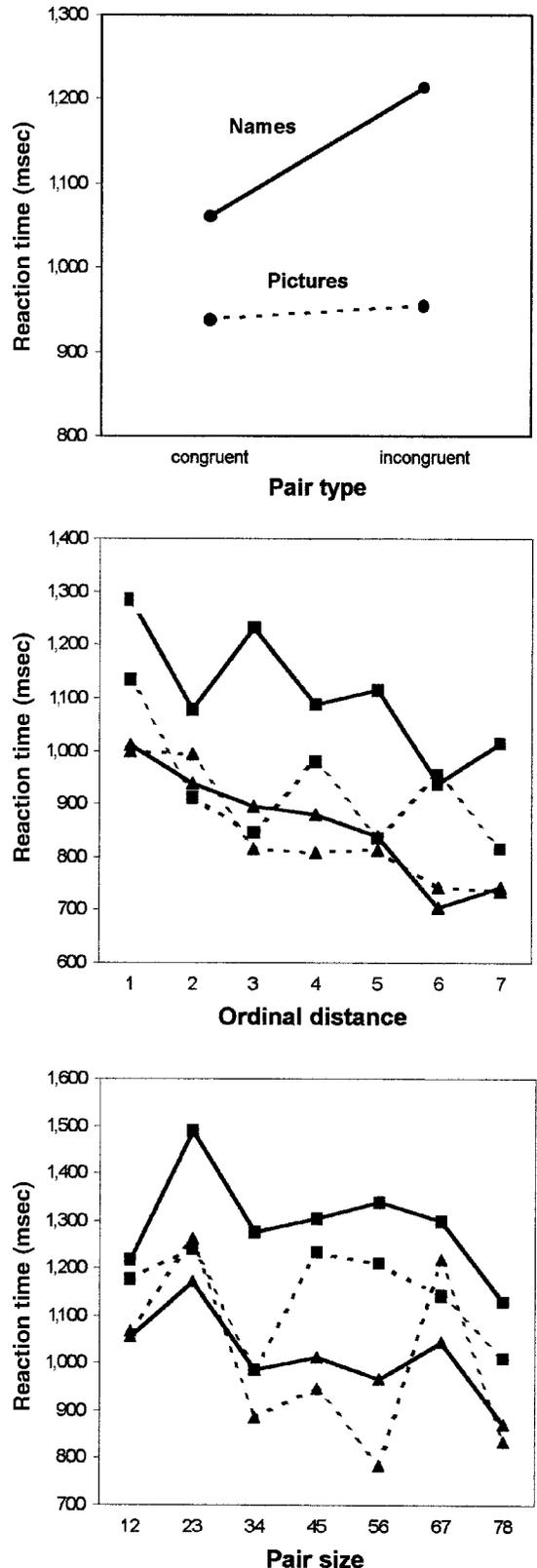


Figure 4. Same as Figure 2, for the data of Experiment 2. Squares: names; triangles: pictures; solid lines: incongruent stimuli; broken lines: congruent stimuli.

9.33, $MS_e = 2,325$, $p < .01$]. The interaction of criterial dimension and pair type (congruent vs. incongruent) underscored the good quality of selective attention to pictures, but its failure for names [$F(1,31) = 103.58$, $MS_e = 1,052$, $p < .01$]. For this larger set of stimuli, too, the participants ignored the words when they were comparing the pictures, but did not ignore the pictures when they were comparing the words. As a result, Stroop interference only plagued the latter comparisons.

In the middle panel of Figure 4, we plotted RT as a function of the ordinal distance separating the two animals in a pair. Separate distance functions are drawn for congruent and incongruent pairs of stimuli for each criterial dimension. Apparent again is the speedier performance with pictures than with words [$F(1,31) = 11.277$, $MS_e = 529,665$, $p < .01$]. The classic distance effect obtained for both congruent and incongruent pairs of stimuli for both dimensions [$F(6,186) = 16.51$, $MS_e = 89,703$, $p < .01$]. For all conditions, RT decreased as the ordinal distance separating members of the pair increased. For names, the large constant separation of the two functions revealed an appreciable Stroop effect; the absence of a vertical separation between the pair of functions for pictures documented the absence of a Stroop effect for comparisons of pictures. The interaction between criterial dimension and pair type [$F(1,31) = 33.59$, $MS_e = 49,135$, $p < .001$] confirmed the results of the visual inspection.

Notable, again, is the additivity of the Stroop and the distance effects. For comparisons of the pictures, a distance effect was present in the data, but a Stroop effect was absent from the same data. For comparisons of the names

in which both effects were present, the effects did not interact ($F < 1$) and both contributed to the outcome in an additive fashion. We conclude that Stroop and distance effects do not derive from the same stage of processing. If the Stroop effect results from processes of stimulus encoding, the distance effect results from a stage other than that of encoding. We will return to this issue.

In the bottom panel of Figure 4, we present the end effect for pairs separated by a single unit of ordinal distance. The data for both pictures and names are complex and involve quadratic components [$F(1,186) = 11.7$, $MS_e = 50,063$, $p < .001$ for names, but $F < 1$ for pictures] apart from the linear ones [$F(1,186) = 7.267$, $MS_e = 50,063$, $p < .01$ for names, and $F(1,186) = 4.959$, $MS_e = 164,136$, $p < .05$ for pictures]. Overall, the end effect entailed RTs that decreased with pair size [$F(6,186) = 18.0$, $MS_e = 81,655$, $p < .01$]. For this subset of the data too, pictures were compared faster than names [$F(1,31) = 5.03$, $MS_e = 1,442,754$, $p < .05$], and a large constant separation between congruent and incongruent pairs was confined to comparisons of the names [$F(1,31) = 5.48$, $MS_e = 122,040$, $p < .05$]. Therefore, a Stroop effect plagued comparisons of names more than it did comparisons of pictures (in fact, it almost never plagued performance with the pictures), although the distinction is somewhat less pronounced in this subset of the data.

The effects of semantic congruity are illustrated in Figure 5 for comparisons of pictures and words, separately for congruent and incongruent stimuli. Despite the variability, several features of the data are discernible. For comparisons of the pictures in Figure 5, the family of curves overlapped; the lack of an RT difference for congruent and incongruent

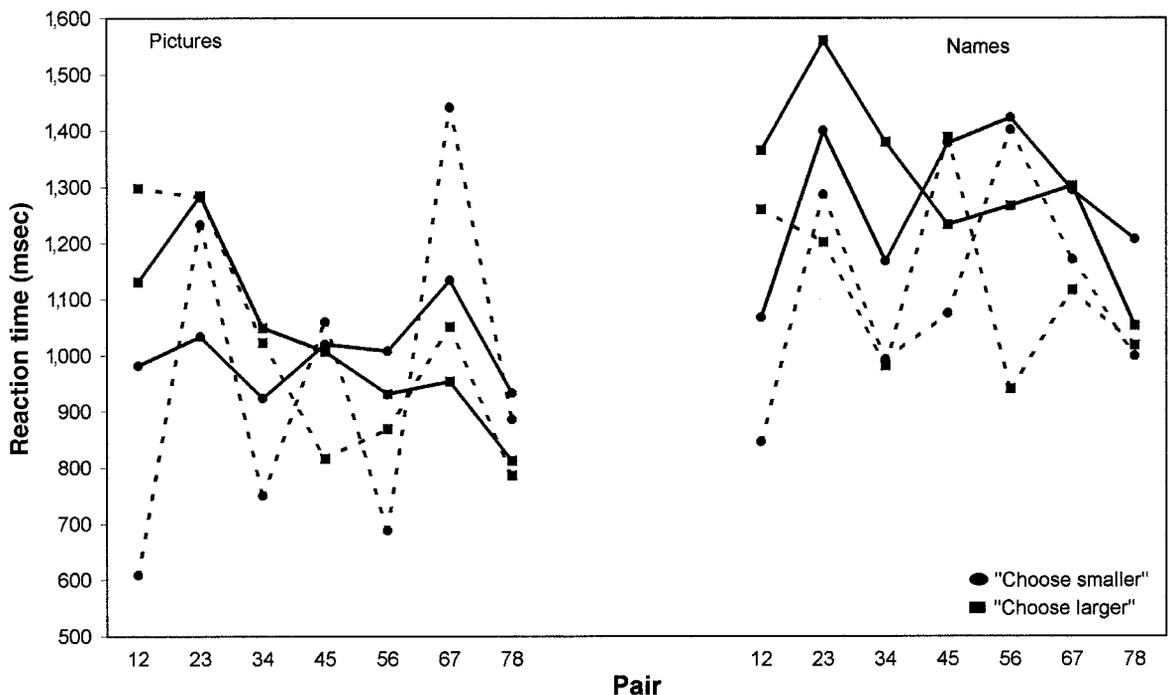


Figure 5. Semantic congruity effects for pictures and names, separately for congruent (broken lines) and incongruent (solid lines) stimuli.

pairs documents the absence of a Stroop effect [$F(1,31) = 2.2, MS_e = 919,429, p > .10$]. By contrast, an appreciable effect of semantic congruity is apparent in the same data: For both congruent and incongruent items, small pairs were responded to faster under the instruction “choose smaller,” and large pairs were responded to faster under the instruction “choose larger” [$F(1,31) = 13.45, MS_e = 122,699, p < .001$]. For comparisons of the names in Figure 5, congruent pairs were responded to faster than incongruent pairs, betraying the presence of a substantial Stroop effect [$F(1,31) = 17.3, MS_e = 425,448, p < .001$]. Equally appreciable effects of semantic congruity, as large indeed as those observed with pictures, were also present [$F(1,31) = 21.08, MS_e = 59,380, p < .001$]. Finally, if the data are compared across pictures and words, the faster general responding to the former is again apparent [$F(1,31) = 7.2, MS_e = 1,983,685, p < .05$].

The interaction of the Stroop effect and semantic congruity is best examined with another, reduced format of the same data in which the test is not burdened by the non-linear components of the curves in Figure 5. These phasic components likely derive from the large variability of the responses inherent in the present design. In Figure 6, we collapsed the data across the four smallest and the four largest pairs. Of interest were the comparisons of names in which both Stroop and semantic congruity effects were present. Salient to the visual inspection of Figure 6 is the parallelism of the two families of functions for names. The semantic congruity effects, expressed by the cross-over interactions, remained invariant across the slowdown (for the incongruent pairs) engendered by the Stroop effect. The visual parallelism, in turn, implies additive effects of Stroop and semantic congruity.

All of the previous effects are visible in the reduced format of Figure 6. For pictures, a negligible Stroop effect of 8.8 msec was accompanied by large effects of semantic congruity for both congruent and incongruent stimuli. For words, by contrast, a large Stroop effect, 169.5 msec, was present along with large semantic congruity effects for congruent and incongruent pairs of stimuli. The interaction of pair size and instruction across both dimensions [$F(1,31) = 40.79, MS_e = 51,473, p < .001$] confirmed the presence of semantic congruity for comparisons of both pictures and names. The interaction of pair type and dimension [$F(1,31) = 11.7, MS_e = 908,004, p < .01$] underscored the presence of the Stroop effect only for comparisons of names. Finally, the absence of the three-way interaction, pair size \times instruction \times pair type ($F < 1$) for names reinforced the additive influence of Stroop and semantic congruity effects.

Discussion

The present results replicate and reinforce the results observed in the first experiment. First, pictures were compared faster than names. Second, the comparisons of pictures were free of intrusions from the irrelevant names, but the comparisons of names suffered Stroop interference

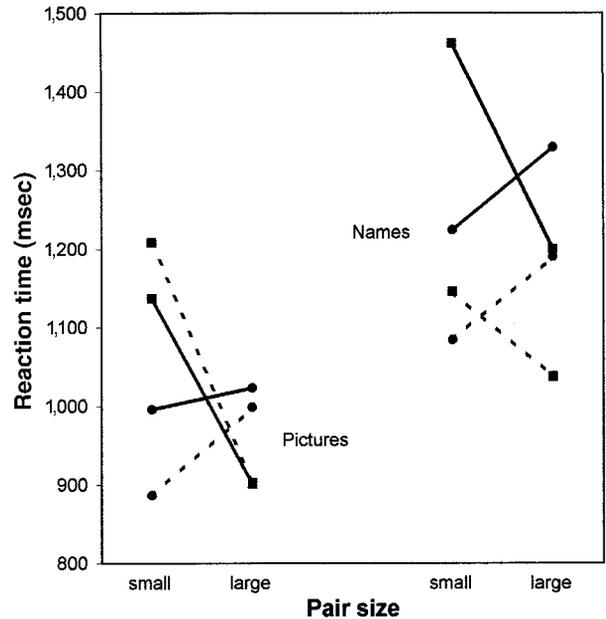


Figure 6. The data of Figure 5 collapsed across the four smallest pairs (small pairs) and the four largest pairs (large pairs).

from the irrelevant pictures. Third, effects of distance, end, and semantic congruity were present in comparisons of both pictures and words. However, none of these effects interacted with the Stroop effect (which, again, was only present in comparisons of names). Of paramount importance was the observed additivity of Stroop and semantic congruity effects.

Regarding the dissociation of Stroop and semantic congruity, the parallelism of the two families of cross-over functions is impressive (see Figure 6). Large and fairly equal effects of semantic congruity are evident for the incongruent stimuli at the top and for the congruent stimuli at the bottom. Their vertical separation depicts a large Stroop effect. Stroop interference amounted to approximately 170 msec in Experiment 2, almost three times its value in Experiment 1. The increase is to be expected, given the larger range of values in Experiment 2. However, what is most revealing about the two phenomena of Stroop and semantic congruity depicted in Figure 6 is their additivity. Moving the bottom functions up a constant duration—that is, removing Stroop interference from the data—aligns them with the functions at the top, betraying a Stroop-invariant effect of semantic congruity.

The demonstration of independence in Experiment 2 provides support for a model in which Stroop and semantic congruity have effects at separate stages of processing. The responses in Experiment 2 were free of confounding by strategic influences. The replication of the results of Experiment 1 thus is impressive. Our interpretation of the additivity of Stroop interference and semantic congruity is that the Stroop effect occurs at an earlier stage of processing than does semantic congruity. The same relation

holds for Stroop and distance effects. The implications of these results for theories of symbolic comparison are discussed next.

GENERAL DISCUSSION

In the present study, we used multidimensional stimuli for mental comparison. The study carries considerable ecological appeal because the stimuli in people's perceptual milieu are multidimensional. In everyday life, people abstract the individual dimension of interest from the complex stimuli in view in order to perform pairwise comparisons of values along the selected dimension. To succeed, people ignore variation along irrelevant dimensions of the same stimuli and focus solely on the target dimension. In the present study, we mimicked this process of selection and comparison by presenting pairs of picture-word compounds, then directing the participant to compare them along either the picture or the word component. Pictures and words are popular stimuli in tasks of symbolic comparison (see Banks & Flora, 1977, and Paivio, 1975, for studies that compare directly the processing of pictures and words in this task). However, in the present study, we combined the pictures and the words into picture-word compounds denoting a common set of referent objects. By the same token, our study imported the Stroop phenomenon into processes of symbolic comparison.

The most important result to emerge from this study is the independence of the Stroop and semantic congruity effects in symbolic comparison. Consequently, the processes that govern (1) input specification of the to-be-compared stimuli and (2) implementation of the comparative instruction that guides the decision are different. The incommensurability of the Stroop and semantic congruity effects carries ramifications for a theoretical resolution.

Re-viewing theories of semantic congruity, it becomes immediately clear that Banks and Root's (1979) *semantic interference model*, a single-stage theory, is incompatible with the present results. In the model, the semantic congruity effect is construed to be a subspecies of the Stroop effect in which, say, the "loudness of the loud stimuli would be an irrelevant attribute that gives rise to an implicit code 'loud' that could interfere with processing of the 'choose quieter' instruction and facilitate the 'choose louder' instruction" (Banks & Root, 1979, p. 140). The model thus captures the semantic congruity effect by assuming that the compared stimuli undergo semantic processing—that is, they are assigned meanings such as "loud" or "soft"—within each act of comparison. These meanings correspond or conflict with those assigned to the comparative instruction—hence the Stroop effect, better known as semantic congruity.

In the present study, a classic version of the Stroop effect was generated and shown to affect performance. Espousing the *semantic interference model*, the to-be-compared stimuli were affected by a pair of Stroop processes. They ought to interact—if indeed semantic congruity is a species of the Stroop phenomenon. However, in contrast to the

model's prediction, the effects of semantic congruity and Stroop did not interact. We reject, perforce, the idea that semantic congruity is a semantic effect engendered by a Stroop process, or the idea that the comparison process is accomplished in a single stage. The *semantic interference model* is untenable under either interpretation. This negative conclusion readily generalizes to all single-stage models entailing semantic processing.

Recently, Leth-Steensen and Marley (2000) have developed a comprehensive computational theory of symbolic comparison. To model semantic congruity, Leth-Steensen and Marley suggested another version of the semantic interference notion. In this instantiation of semantic interference, the comparative used (e.g., "Larger?") activates the appropriate instructional pathway and deactivates the complementary instructional pathway ("Smaller?"). However, the selectivity of activation is not perfect. For any given comparison, the irrelevant instructional pathway is also activated and competes with the processing within the relevant pathway. Following Banks and Root's (1979) original idea, Leth-Steensen and Marley further assumed that the stimuli themselves undergo semantic analysis within each comparison. The respective end anchors—already associated cognitively with "largeness" and "smallness"—confer meaning on the stimuli by virtue of similarity. The meaning of the stimuli then determines the extent of the competition between the two instructional pathways. The difference in activation favoring the relevant pathway is magnified when the meaning of the stimuli is congruent with that of the instruction; otherwise, the difference is attenuated. These variations result in differential rates of evidence accrual (hence different RTs) for a given pair of stimuli under the two forms of the comparative instruction. In this computational rendition, too, semantic congruity is rooted in a Stroop process involving the respective meanings of the stimuli and the instructions.

There appears to be two difficulties with Leth-Steensen and Marley's (2000) interpretation of semantic congruity, one logical, the other empirical. For the former, it is not *prima facie* clear why an instruction other than that posed experimentally is activated on a trial (but see Wenger, 1994, for a general approach centered on the notion of automatic activation of task-irrelevant instructions in all situations). The assumption is also inconsistent with the common finding (Marschark & Paivio, 1979, 1981) that semantic congruity is larger in tasks in which the form of the comparative instruction is manipulated in blocks of trials than in tasks entailing a random trial-by-trial manipulation. Considering the present results, Leth-Steensen and Marley's interpretation of semantic congruity is suspect given the independence of Stroop and semantic congruity effects. The observed independence is unlikely to occur under the hypothesis of semantic interference engendered by the stimuli.

Model diagnosis was further assisted by breaking down the Stroop-incongruent pairs into two categories, strongly and weakly incongruent stimuli (SI and WI, respectively).³ The former category included those incongruent pairs in

which the pictures and the words belonged in different-size categories (e.g., the words *fly* and *cat* superimposed on the respective pictures of elephant and cow). The latter category included those incongruent pairs in which the pictures and the words belonged in same-sized categories (e.g., the words *fly* and *bird* superimposed on the respective pictures of butterfly and dog). According to the Leth-Steensen and Marley (2000) view, when the activation from the stimulus code pathways is weakened, the competition between the instructional pathways will be weakened. As a result, semantic congruity should be reduced for SI pairs compared with WI pairs (Figure 7).

The displays in Figure 7 show that semantic congruity did *not* differ for SI and WI pairs in either experiment. In Experiment 1, SI stimuli were compared more slowly than WI stimuli by 30 msec on average [a subspecies of the Stroop effect; $F(1,19) = 12.41, MS_e = 708.9, p < .001$]. Nevertheless, semantic congruity remained invariant as the magnitude of Stroop interference (SI vs. WI) did not interact with semantic congruity ($F < 1$). In Experiment 2, the Stroop slowdown for SI pairs was 47.6 msec [$F(1,31) = 16.15, MS_e = 2,570, p < .001$]. Of most importance, again, was the lack of an interaction between Stroop incongruity and semantic congruity ($F < 1$). Contrary to the prediction based on the Leth-Steensen and Marley (2000) model, semantic congruity remained invariant in the face of differing amounts of competition between the instructional pathways.

Another partition of the incongruent stimuli may serve to test the dissociation of the Stroop and semantic congruity effects.⁴ Recall that the incongruent pairs in this study were defined in terms of the composition of their members: At least one member of the pair consisted of a conflicting picture–word stimulus. However, another source of conflict exists at the level of the comparative *response*: The response to the target dimension (name) may or may not conflict with that implied by the distractor dimension (picture). Consider the words *cat* and *cow*. In one combination, *cat* is embedded in a picture of a fly, and *cow* is embedded in a picture of an elephant. Although the pair is incongruent (both member stimuli comprise conflicting verbal and visual components), the verbal and visual components converge on the same response (cow is larger than cat, and elephant is larger than fly). In another combination, *cat* is embedded in a picture of an elephant, and *cow* in a picture of a fly. The latter pair is incongruent with respect to both member composition *and* the responses to the verbal and visual components. Let us call the first class of incongruent pairs semantically incongruent (SEI) and the second class semantically and response incongruent (SEREI). We tested the relationship between the effects of Stroop and semantic congruity separately for the two types of Stroop-incongruent stimuli.

SEREI stimuli were responded to 892.3 msec on average, and SEI stimuli were responded to 866.1 msec on average [$t(19) = 3.58, p < .002$]. The 26.2-msec difference is the additional toll on performance claimed by stimuli affected by two sources of Stroop conflict as compared with

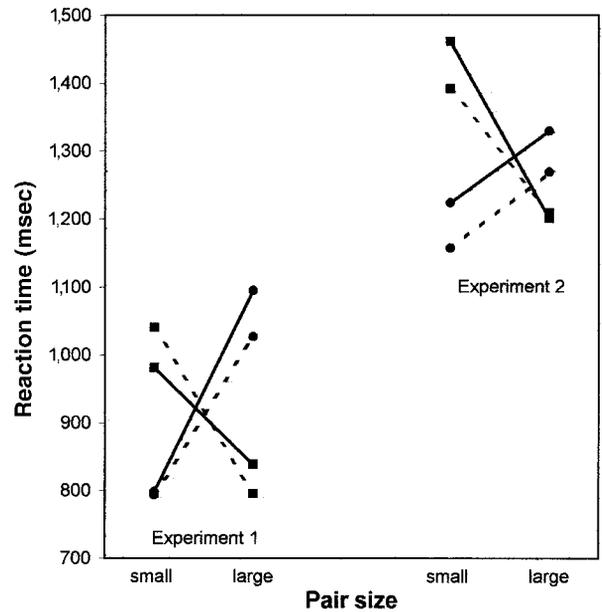


Figure 7. Semantic congruity effects for picture–word compounds entailing strong (solid lines) and weak (broken lines) conflict between components. Effects of semantic congruity for names are depicted for the two types of Stroop-incongruent compounds in Experiment 1 (left side) and Experiment 2 (right side).

a single source. We calculated the interaction of Stroop and semantic congruity, with the former effect defined alternatively as SEI or SEREI. Neither interaction was present in the data [$F(2,28) = 1.152, MS_e = 223,563, p > .05$ for SEI and $F(2,28) = 1.747, MS_e = 268,843, p > .05$ for SEREI]. Hence, Stroop interference did not interact with semantic congruity regardless of the species of the interference, whether double or single Stroop. The failure to find an interaction between Stroop interference, overall, and semantic congruity cannot be explained as being due to competing effects of semantic and response congruence.

Nevertheless, response interference did make a difference: The semantic congruity effect was smaller for SEREI than for SEI [$F(2,38) = 11.841, MS_e = 110,597, p < .001$]. Taken alone, neither effect of semantic congruity interacted with its respective Stroop effect (with the incongruent pairs defined in turn as SEI and as SEREI). Taken together, however, response congruity affected semantic congruity. The two sources of Stroop interference should be carefully disentangled in future research (Shaki, 2001). Incidentally, the two sources can be discerned only in tasks of comparison; they are indistinguishable in the classic Stroop task that concerns only single stimuli.

One way to salvage some of the models, we alluded, is to ascribe the Stroop and semantic congruity effects to different stages of processing; hence, the lack of interaction. According to Banks's (1977) *discrete code model*, a very influential theory, the congruity effect is located in a matching stage and follows from the transformation of the semantic codes of the stimuli when they do

not match that of the instructions. The Stroop effect (when it applies) is located in the encoding stage in which “a set of processing functions . . . generates a semantic code for the stimulus” (Banks & Flora, 1977, p. 288). The size-congruity effect (an interaction between the physical size of the symbol and that of its reference, Paivio, 1975) or the interaction between the words *up* and *down* and their relative locations (Clark & Brownell, 1975)—subspecies of the generic Stroop effect—both “point to the encoding stage. Thus, . . . [Stroop] congruity facilitates encoding . . . and [Stroop] incongruity interferes with encoding” (Banks & Flora, 1977, p. 289).

We agree that Stroop processes function in the encoding stage. They are instrumental in specifying the mental magnitudes of the to-be-compared stimuli. Without prior valuation of the pertinent stimuli, no meaningful comparison between them is possible. The Stroop effect taps the retardation of valuation if an irrelevant attribute conflicts with the task-relevant attribute; it also taps the speeding up of valuation when an irrelevant attribute corresponds to the task-relevant one. Nevertheless, one should recall that Stroop processes are based on the activation of meaning (the very classification congruent vs. incongruent stimuli is semantically based). Theories that construe semantic congruity to involve changes of stimulus meaning (Banks, 1977; Cech & Shoben, 1985; Shoben, Cech, Schwanenflugel, & Sailor, 1989) and, certainly, theories that construe semantic congruity to comprise another instance of a Stroop effect (Banks & Root, 1979; Leth-Steensen & Marley, 2000) must envisage a rigorous segregation of processing stages. Otherwise, it is difficult to understand how two Stroop processes, or, at the least, a pair of processes that share in semantic activation, are independent, unless, of course, semantic congruity does not entail appreciable processes of semantic analysis (its name notwithstanding), a possibility entertained in evidence accrual models of stimulus comparison.

Our results thus accord best with theories that (1) do not attribute semantic congruity to semantic (re)analysis of the stimuli, and that (2) place the effect within the decision process. These conclusions are supported (though, by no means, proved) by (1) the lack of interaction between Stroop and semantic congruity effects, (2) the semantic nature of the Stroop phenomenon itself, and (3) the evidence that Stroop effects occur at the encoding stage. Taken as a whole, our results fit best with Petrusic’s (1992) evidence accrual model. In Petrusic’s model, semantic congruity is part and parcel of the comparative decision process. Petrusic (see also Petrusic & Baranski, 1989) has shown the effect to be sensitive to speed-accuracy relations at different levels of processing and concluded that the effect occurs at each discrete evidence accrual event within a global evidence-accrual-based process. Random-walk evidence accrual models (Birnbaum & Jou, 1990; Link, 1990) attribute the effect to differential starting points of the walk dictated by the instructions (typically implemented through bias parameters). The present results are commensurate with this family of models. How-

ever, in fairness to investigators who endorse versions of a multistage semantic model, such models are not rejected by our data—as long as they maintain a strict separation of processing stages. The plausibility (and testing) of strict serial processing is beyond the purview of the present study.

Our results show that the Petrusic (1992) model must also entail a separate stage of encoding that includes Stroop effects. If Stroop interference occurs at each pass through the accrual-based decision, the larger the Stroop effect, the larger the effect of semantic congruity. According to the theory developed by Petrusic, any process that slows the evidence accrual process will result in enhanced semantic congruity. Our data show, instead, semantic congruity to be independent of the magnitude of Stroop interference. Therefore, we conclude that the accrual process begins only after the stimulus pair is fully encoded.

The present data show the distance effect, too, to be independent of the Stroop effect. As a result, we conclude that the distance effect also occurs *after* encoding. Several possibilities can be considered regarding its locus. Analogue theories (e.g., Holyoak, 1978; Jamieson & Petrusic, 1975) and particularly random-walk models (Birnbaum & Jou, 1990; Link, 1990, 1992) easily accommodate the distance effect as an integral part of the comparative decision process. Incidentally, Duncan and McFarland (1980) have reached the same conclusion on the basis of additive-factors analysis. Petrusic’s (1992) model is less explicit regarding the locus of the distance effect, but, on the basis of its underlying logic, the distance effect, too, is part of decision processing. Alternatively, the distance effect might occur prior to decision, although after the elementary processes of stimulus perception (Banks, 1977). Banks identified a series of initial stages of processing *collectively* referred to as encoding. The distance effect occurs in the final stage of a global encoding operation.

We conclude this discussion with two observations regarding the Stroop effect itself. First, contrary to the widespread belief that it is ubiquitous and automatic (but see recent research by Besner, 2001; Besner, Stolz, & Boutilier, 1997; Stolz & Besner, 1999), we showed the Stroop effect to be selective and optional. It affected performance with words, but the effect was conspicuously absent from that with pictures. Pictures intruded on comparison performance with the words, but the words did not affect the comparisons of pictures. The present results join those previously obtained in our laboratory (e.g., Algom, Dekel, & Pansky, 1996; Ariei & Algom, 1997, 2002; Dishon-Berkovits & Algom, 2000; Pansky & Algom, 1999, in press; Shalev & Algom, 2000; see also Melara & Marks, 1990; Melara & Mounts, 1993; Sabri, Melara, & Algom, 2001) showing the magnitude, direction, and very presence of the Stroop effect to depend on a few factors of context. One factor is relative dimensional discriminability, the ease with which the respective stimulus values are perceived and discriminated along each of the constituent dimensions. In the present study, the pictures constituted the more discriminable or easier dimension, and, as a result, they interfered with performance with the words, but not

vice versa. Incidentally, a review of the vast literature on the original Stroop effect reveals that, as a rule, the words used were much more discriminable than the colors used; this asymmetry itself might have produced the classic Stroop asymmetry in interference. Indeed, when the words and the colors were matched in advance on discriminability, the Stroop effect collapsed; and when the colors were made to be more discriminable than the words, the reverse of the classic Stroop effect obtained (Melara & Mounts, 1993).

Finally, the direction of the Stroop interference obtained—pictures intruding on performance with words, but not vice versa—uncovers the underlying nature of mental comparison as a categorization task. The asymmetric interference is consistent with picture superiority in symbolic comparisons when pictures and words are presented separately (Banks & Flora, 1977; Paivio, 1975). Both findings are commensurate with the universal advantage of pictures in tasks involving semantic processing—notably, categorization—with single stimuli (Smith & Magee, 1980). Pictures (unlike words) engage the semantic system directly—hence their relative advantage in tasks involving semantic processing. The present results suggest that the common cognitive task of mental comparison, too, is semantic in nature, likely involving processes of categorization. The comparative instructions (“Larger?” or “Smaller?”) induce processes of categorization even in the case in which the comparison involves physical stimuli. Banks’s (1977; Banks & Flora, 1977; Banks, Mermelstein, & Yu, 1982; Banks & Root, 1979) pioneering studies have tapped the semantic nature of the mental comparison task. The truly surprising message of our study is that semantic congruity—the nominal candidate for capturing the semantic influences in this task—may entail less semantic processing than would have been expected.

REFERENCES

- ALGOM, D., DEKEL, A., & PANSKY, A. (1996). The perception of number from the separability of the stimulus: The Stroop effect revisited. *Memory & Cognition*, *24*, 557-572.
- ARIEH, Y., & ALGOM, D. (1997). [Reading words in Stroop-like tasks: The effect of the contingency between semantic and visual components]. *Psychologia*, *6*, 7-19 [in Hebrew].
- ARIEH, Y., & ALGOM, D. (2002). Processing picture-word stimuli: The contingent nature of picture- and of word-superiority. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 221-232.
- BANKS, W. P. (1977). Encoding and processing of symbolic information in comparative judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11, pp. 101-159). New York: Academic Press.
- BANKS, W. P., CLARK, H. H., & LUCY, P. (1975). The locus of the semantic congruity effect in comparative judgments. *Journal of Experimental Psychology: Human Perception & Performance*, *104*, 35-47.
- BANKS, W. P., & FLORA, J. (1977). Semantic and perceptual processes in symbolic comparisons. *Journal of Experimental Psychology: Human Perception & Performance*, *3*, 278-290.
- BANKS, W. P., MERMELSTEIN, R., & YU, H. K. (1982). Discriminations among perceptual and symbolic stimuli. *Memory & Cognition*, *10*, 265-278.
- BANKS, W. P., & ROOT, M. (1979). Semantic congruity effects in judgments of loudness. *Perception & Psychophysics*, *26*, 133-142.
- BANKS, W. P., WHITE, H., STURGILL, W., & MERMELSTEIN, R. (1983). Semantic congruity and expectancy in symbolic judgment. *Journal of Experimental Psychology: Human Perception & Performance*, *9*, 560-582.
- BESNER, D. (2001). The myth of ballistic processing: Evidence from Stroop’s paradigm. *Psychonomic Bulletin & Review*, *8*, 324-330.
- BESNER, D., STOLZ, J. A., & BOUTILIER, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, *4*, 221-225.
- BIRNBAUM, M. H., & JOU, W., JR. (1990). A theory of comparative response times and “difference” judgments. *Cognitive Psychology*, *22*, 184-210.
- CECH, C. G. (1995). Is congruity due to encoding? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *5*, 1275-1288.
- CECH, C. G., & SHOEN, E. J. (1985). Context effects in symbolic magnitude comparisons. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 299-315.
- CECH, C. G., SHOEN, E. J., & LOVE, M. (1990). Multiple congruity effects in judgments of magnitude. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 1142-1152.
- CLARK, H. H., & BROWNELL, H. H. (1975). Judging up and down. *Journal of Experimental Psychology: Human Perception & Performance*, *1*, 339-352.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- DISHON-BERKOVITS, M., & ALGOM, D. (2000). The Stroop effect: It is not the robust phenomenon that you have thought it to be. *Memory & Cognition*, *28*, 1437-1449.
- DUNCAN, E. M., & MCFARLAND, C. E., JR. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory & Cognition*, *8*, 612-622.
- FARELL, B. (1985). “Same-different” judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, *98*, 419-456.
- FAUL, F., & ERDFELDER, E. (1992). *GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS* [Computer program]. Bonn: Bonn University, Department of Psychology.
- HOLYOAK, J. K. (1978). Comparative judgments with numerical reference points. *Cognitive Psychology*, *10*, 203-243.
- JAMIESON, D. G., & PETRUSIC, W. M. (1975). Relational judgments with remembered stimuli. *Perception & Psychophysics*, *18*, 373-378.
- KEPPEL, G. (1973). *Design and analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- LETH-STEENSEN, C., & MARLEY, A. A. J. (2000). A model of response time effects in symbolic comparison. *Psychological Review*, *107*, 62-100.
- LINK, S. W. (1990). Modeling imageless thought: The relative judgment theory of numerical comparisons. *Journal of Mathematical Psychology*, *34*, 2-41.
- LINK, S. W. (1992). *The wave theory of difference and similarity*. Hove, U.K.: Erlbaum.
- MARKS, D. F. (1972). Relative judgment: A phenomenon and a theory. *Perception & Psychophysics*, *11*, 156-160.
- MARSCHARK, M., & PAIVIO, A. (1979). Semantic congruity and lexical marking in symbolic comparisons: An expectancy hypothesis. *Memory & Cognition*, *7*, 175-184.
- MARSCHARK, M., & PAIVIO, A. (1981). Congruity and the perceptual comparison task. *Journal of Experimental Psychology: Human Perception & Performance*, *7*, 290-308.
- MELARA, R. D., & MARKS, L. E. (1990). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory & Cognition*, *18*, 477-495.
- MELARA, R. D., & MOUNTS, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition*, *21*, 627-645.
- MOYER, R. S., & DUMAIS, S. T. (1978). Mental comparison. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12, pp. 117-155). New York: Academic Press.
- MOYER, R. S., & LANDAUER, T. K. (1967). The time required for judgment of numerical inequality. *Nature*, *215*, 1519-1520.
- PAIVIO, A. (1975). Perceptual comparisons through the mind’s eye. *Memory & Cognition*, *3*, 635-647.
- PANSKY, A., & ALGOM, D. (1999). Stroop and Garner effects in comparative judgment of numerals: The role of attention. *Journal of Experimental Psychology: Human Perception & Performance*, *25*, 39-58.
- PANSKY, A., & ALGOM, D. (in press). Comparative judgment of nu-

- merosity and numerical magnitude: Attention preempts automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.
- PETRUSIC, W. M. (1992). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 962-986.
- PETRUSIC, W. M., & BARANSKI, J. V. (1989). Semantic congruity effects in perceptual comparisons. *Perception & Psychophysics*, **45**, 439-452.
- SABRI, M., MELARA, R. D., & ALGOM, D. (2001). A confluence of contexts: Asymmetric versus global failure of selective attention to Stroop dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 515-537.
- SHAKI, S. (2001). *Selective attention in classification and in comparison: The role of task*. Unpublished doctoral dissertation, Tel-Aviv University.
- SHALEV, L., & ALGOM, D. (2000). Stroop and Garner effects in and out of Posner's beam: Reconciling two conceptions of selective attention. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 997-1017.
- SHOBEN, E. J., CECHE, C. G., SCHWANENFLUGEL, P. J., & SAILOR, K. M. (1989). Serial position effects in comparative judgments. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 273-286.
- SHOBEN, E. J., SAILOR, K. M., & WANG, M.-Y. (1989). The role of expectancy in comparative judgments. *Memory & Cognition*, **17**, 18-26.
- SMITH, M. C., & MAGEE, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General*, **109**, 373-392.
- STERNBERG, S. (1969). The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, **30**, 276-315.
- STOLZ, J. A., & BESNER, D. (1999). On the myth of automatic semantic activation in reading. *Current Directions in Psychological Science*, **8**, 61-65.
- WENGER, D. M. (1994). Ironic processes of mental comparison. *Psychological Review*, **101**, 34-52.
- WINER, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.

NOTES

1. There is no consensual method in the literature for calculating the magnitude of the semantic congruity effect. Therefore, the values reported should be treated with caution. Of consequence are the results of the statistical analyses involving semantic congruity.

2. For accuracy, the participants made fewer errors with pictures (1.89%) than with names (2.83%), but the difference was not statistically significant [$t(38) = -1.399, p > .05$]. Stroop effects emerged only for names, but the difference was not statistically significant [$t(19) = -1.325, p > .05$]. The semantic congruity effect for errors mirrored the pattern obtained for RT, but the effects were statistically significant for neither pictures nor names [$F < 1$, and $F(1,19) = 3.09, MS_e = 32.26, p > 0.05$, respectively].

3. We thank an anonymous reviewer for suggesting this analysis.

4. We thank a second anonymous reviewer for suggesting the analysis. We made the calculations on the results of Experiment 1 in which the design entailed equal numbers of observations in the various cells.

(Manuscript received June 26, 2000;
revision accepted for publication June 1, 2001.)